wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 09:59 AM GMT

PDB ID	:	3K7A
Title	:	Crystal Structure of an RNA polymerase II-TFIIB complex
Authors	:	Liu, X.; Bushnell, D.A.; Wang, D.; Calero, G.; Kornberg, R.D.
Deposited on	:	2009-10-12
Resolution	:	3.80 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

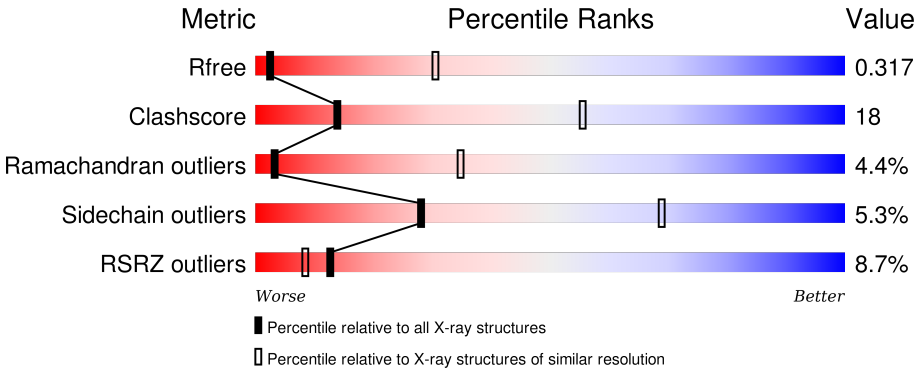
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



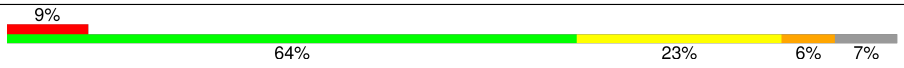
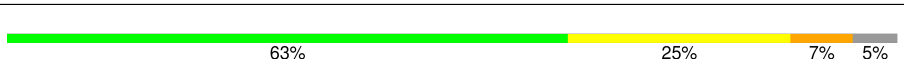
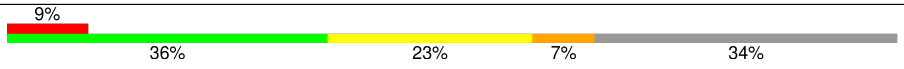
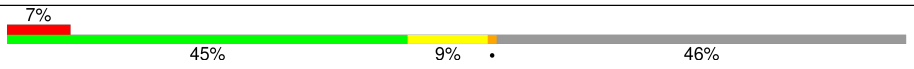
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1317 (4.10-3.50)
Clashscore	102246	1458 (4.10-3.50)
Ramachandran outliers	100387	1397 (4.10-3.50)
Sidechain outliers	100360	1392 (4.10-3.50)
RSRZ outliers	91569	1325 (4.10-3.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1733	4% 49% 29% • 19%
2	B	1224	13% 53% 33% • 8%
3	C	318	55% 27% • 16%
4	E	215	9% 79% 21%
5	F	155	34% 18% • 46%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	H	146	
7	I	122	
8	J	70	
9	K	120	
10	L	70	
11	M	345	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
12	ZN	M	346	-	-	-	X

2 Entry composition

There are 12 unique types of molecules in this entry. The entry contains 29029 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1408	11052	6966	1936	2089	61	0	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1122	8845	5595	1551	1644	55	0	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	266	2095	1317	348	417	13	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	214	1752	1111	309	321	11	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	84	679	434	115	127	3	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	H	133	1068	673	180	211	4	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	I	118	Total	C	N	O	S	0	0	0
			967	594	178	185	10			

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	J	65	Total	C	N	O	S	0	0	0
			532	339	93	94	6			

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	K	114	Total	C	N	O	S	0	0	0
			919	590	156	171	2			

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	L	46	Total	C	N	O	S	0	0	0
			363	224	72	63	4			

- Molecule 11 is a protein called Transcription initiation factor IIB.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
11	M	187	Total	C	N	O	0	0	0
			748	374	187	187			

- Molecule 12 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
12	J	1	Total	Zn	0	0
			1	1		
12	B	1	Total	Zn	0	0
			1	1		
12	I	2	Total	Zn	0	0
			2	2		
12	C	1	Total	Zn	0	0
			1	1		
12	A	2	Total	Zn	0	0
			2	2		

Continued on next page...

Continued from previous page...

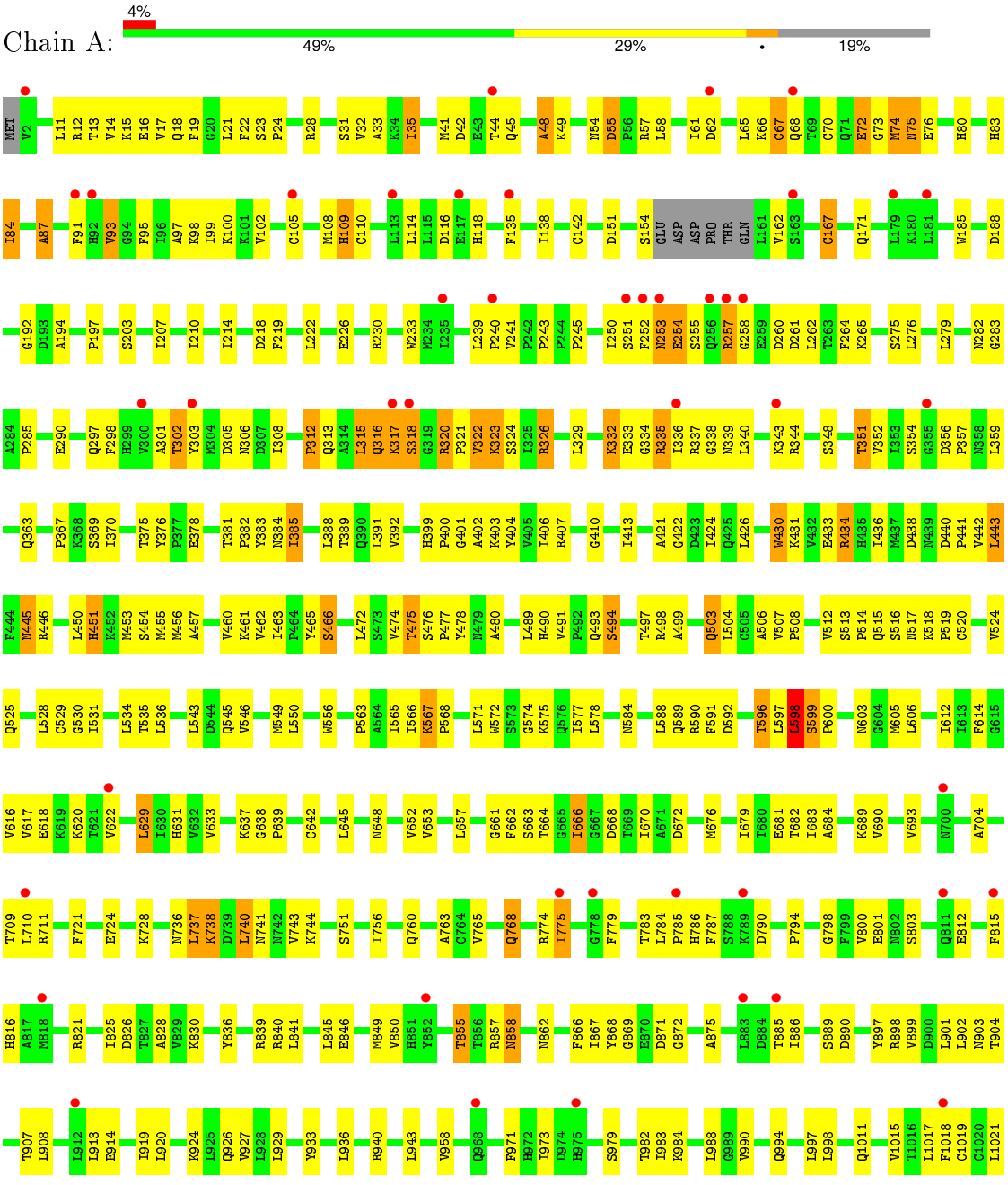
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
12	L	1	Total Zn 1 1	0	0
12	M	1	Total Zn 1 1	0	0

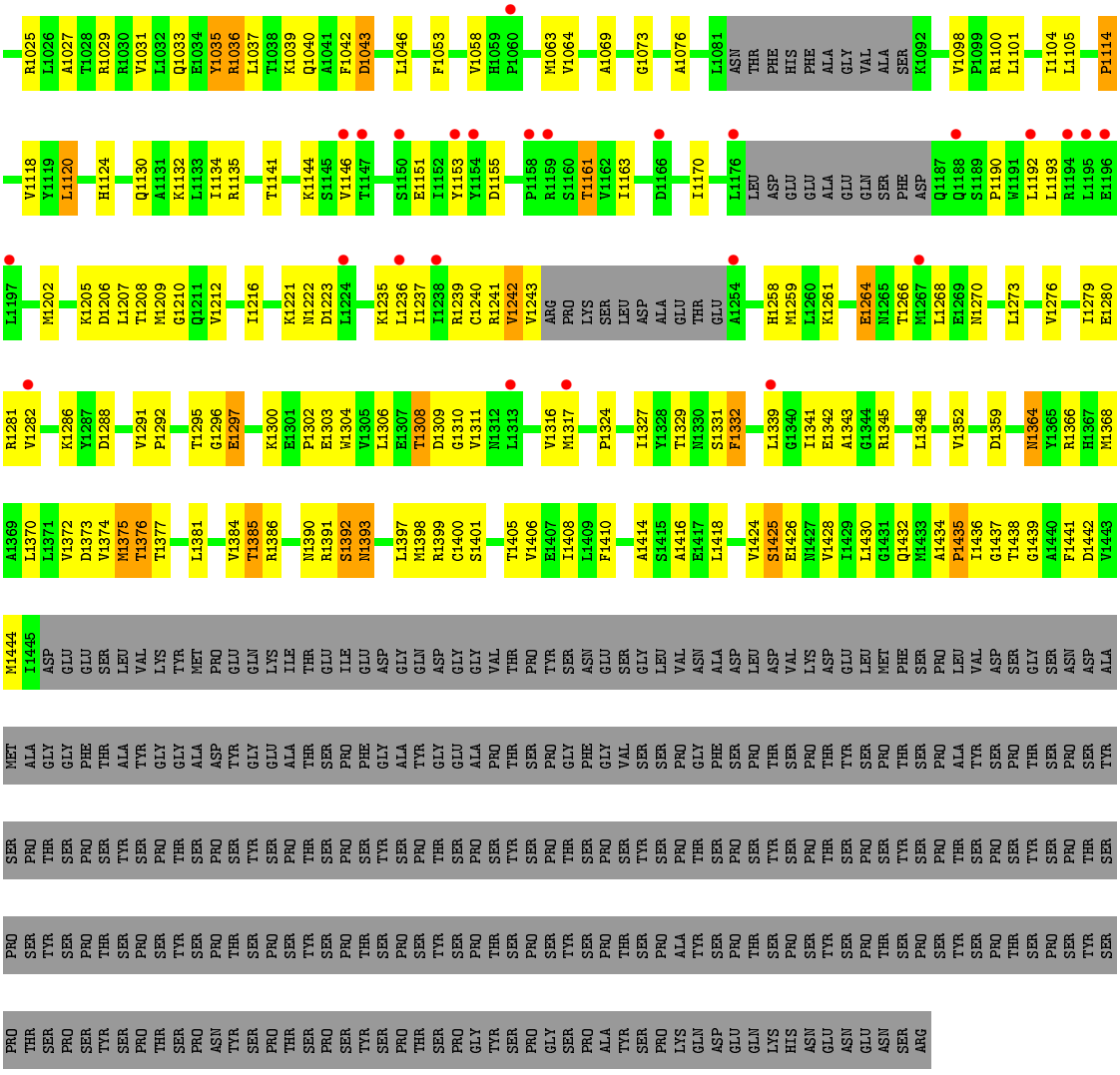
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: DNA-directed RNA polymerase II subunit RPB1

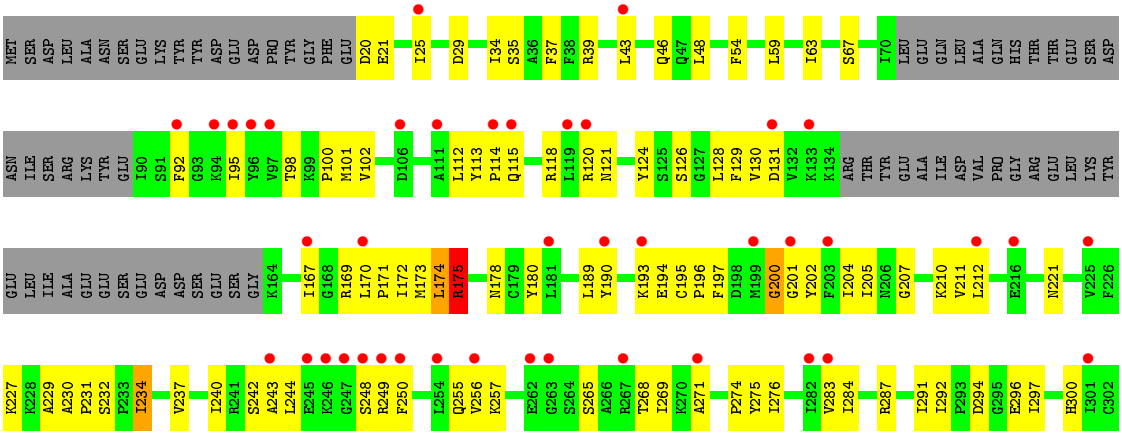
Chain A:

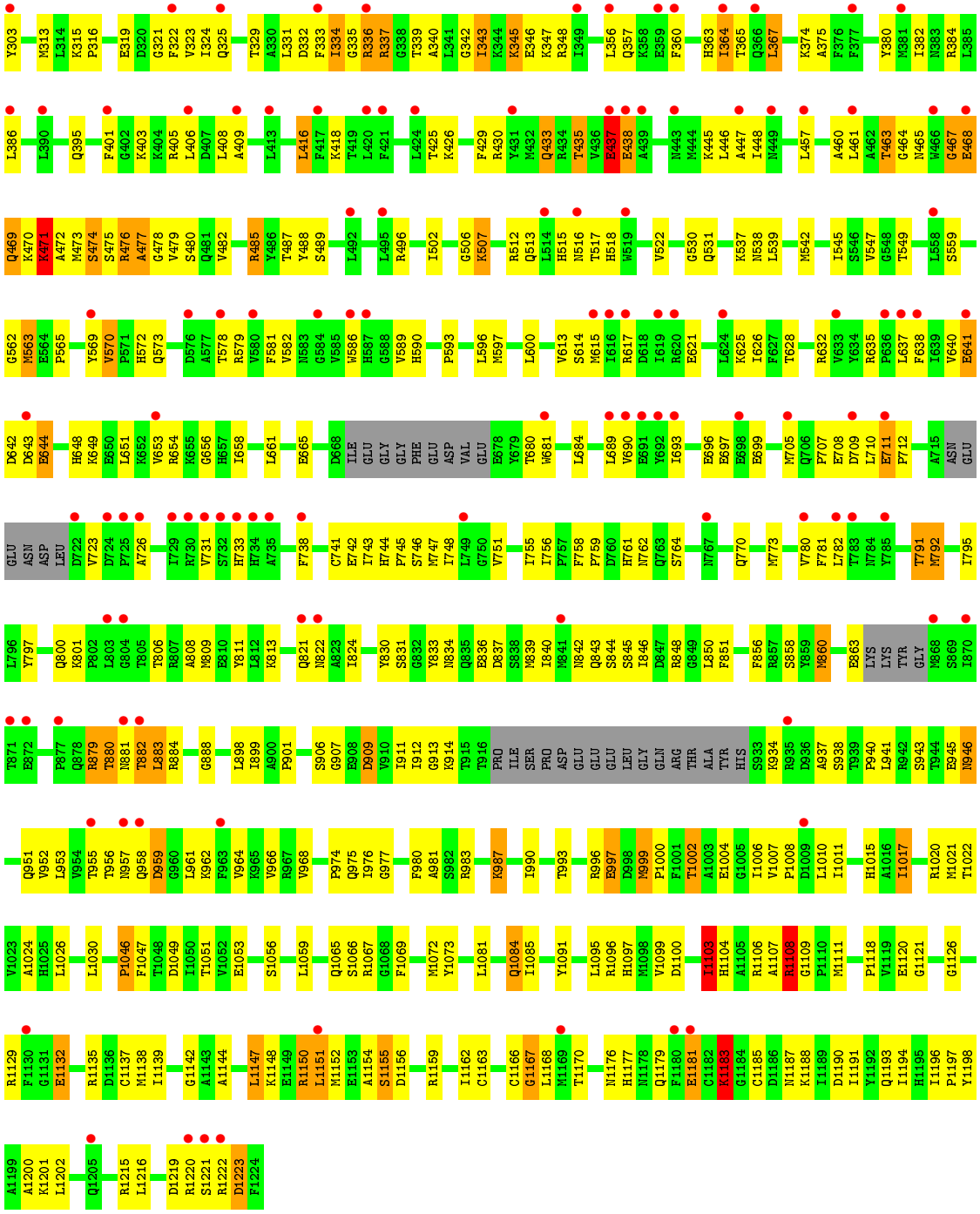




• Molecule 2: DNA-directed RNA polymerase II subunit RPB2

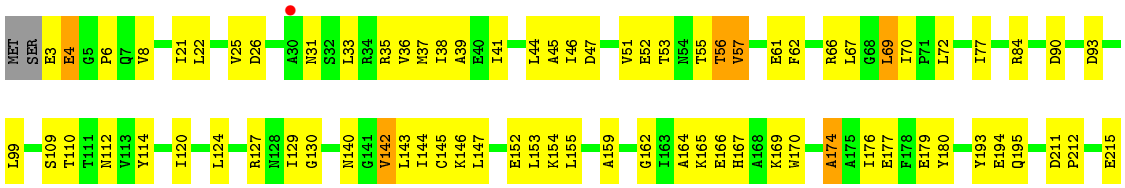
Chain B:

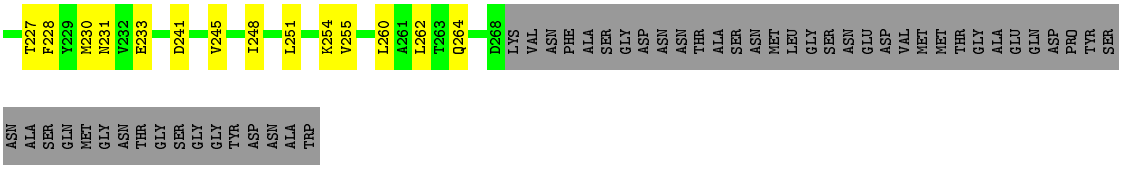




• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

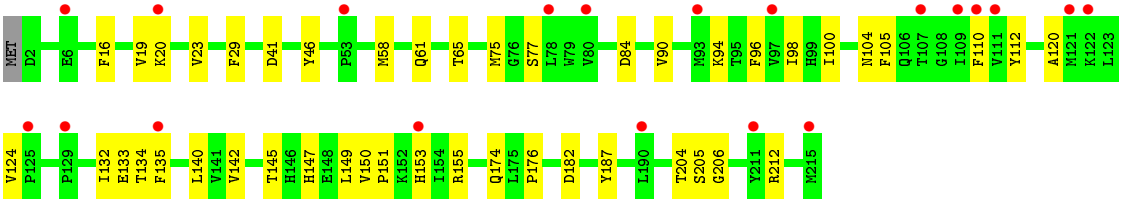
Chain C:





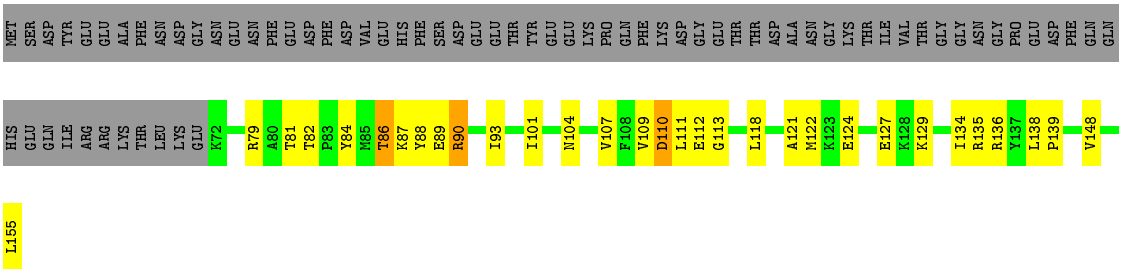
• Molecule 4: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E:



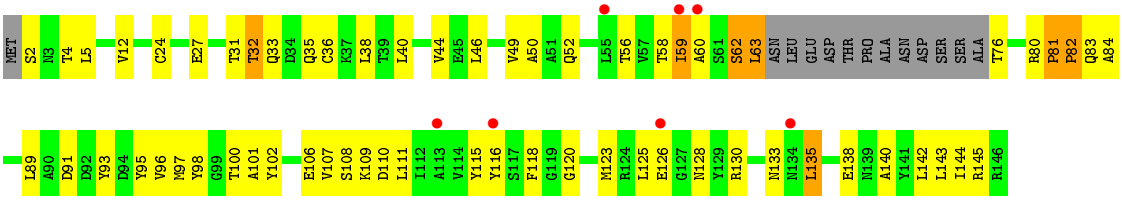
• Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:



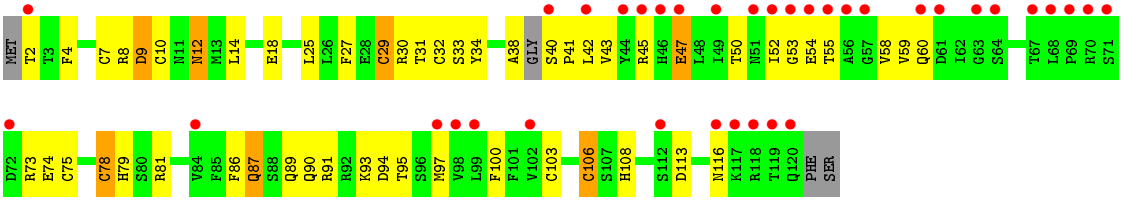
• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:



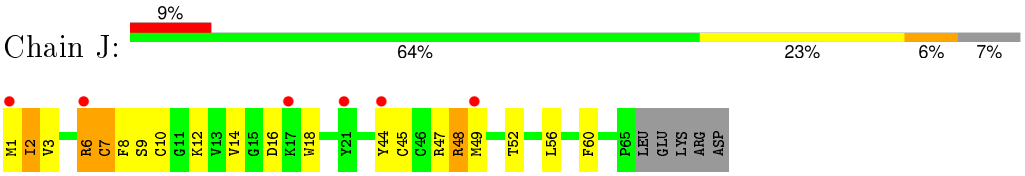
• Molecule 7: DNA-directed RNA polymerase II subunit RPB9

Chain I:



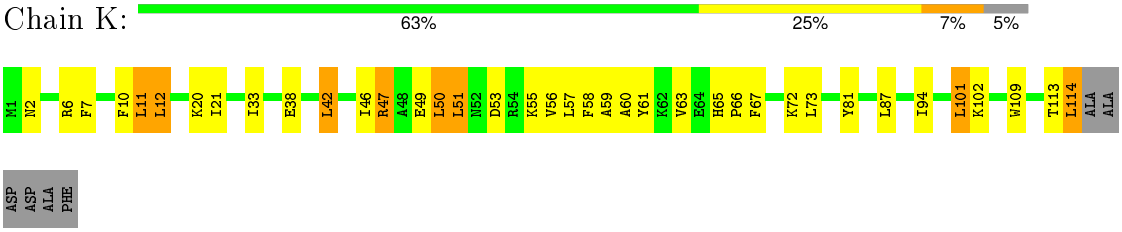
• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:



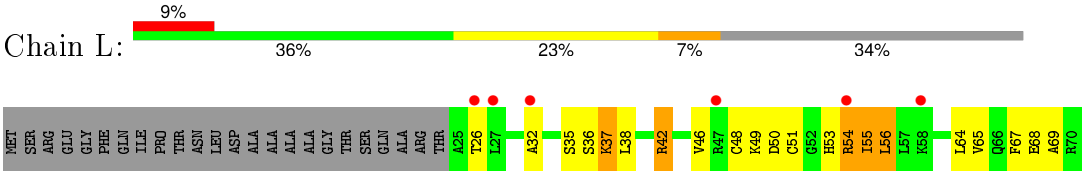
- Molecule 9: DNA-directed RNA polymerase II subunit RPB11

Chain K:



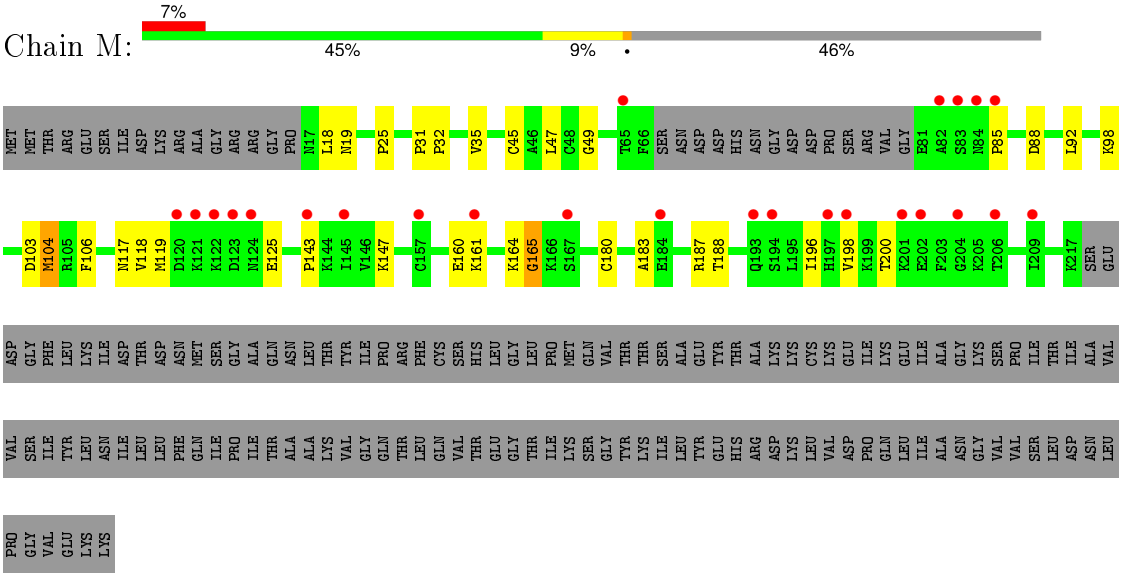
- Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L:



- Molecule 11: Transcription initiation factor IIB

Chain M:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	204.15Å 216.21Å 420.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.80 148.44 – 3.80	Depositor EDS
% Data completeness (in resolution range)	95.5 (50.00-3.80) 95.5 (148.44-3.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	1.55 (at 3.78Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.263 , 0.313 0.265 , 0.317	Depositor DCC
R_{free} test set	4320 reflections (4.94%)	DCC
Wilson B-factor (Å^2)	126.3	Xtriage
Anisotropy	0.598	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 152.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Outliers	0 of 168084 reflections	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	29029	wwPDB-VP
Average B, all atoms (Å^2)	188.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.98% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.52	0/11248	0.67	0/15211
2	B	0.50	0/9016	0.65	0/12165
3	C	0.47	0/2133	0.64	0/2891
4	E	0.41	0/1788	0.56	0/2406
5	F	0.61	0/691	0.72	0/933
6	H	0.45	0/1086	0.61	0/1470
7	I	0.44	0/984	0.61	0/1323
8	J	0.50	0/541	0.66	0/727
9	K	0.55	0/937	0.68	0/1265
10	L	0.52	0/365	0.78	0/485
11	M	0.34	0/746	0.80	2/929 (0.2%)
All	All	0.50	0/29535	0.66	2/39805 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
11	M	165	GLY	N-CA-C	-5.23	100.03	113.10
11	M	161	LYS	N-CA-C	5.08	124.73	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	11052	0	11130	470	0
2	B	8845	0	8816	373	0
3	C	2095	0	2051	71	0
4	E	1752	0	1776	33	0
5	F	679	0	701	24	0
6	H	1068	0	1040	51	0
7	I	967	0	925	45	0
8	J	532	0	542	32	0
9	K	919	0	929	46	0
10	L	363	0	386	12	0
11	M	748	0	200	22	0
12	A	2	0	0	0	0
12	B	1	0	0	0	0
12	C	1	0	0	0	0
12	I	2	0	0	0	0
12	J	1	0	0	0	0
12	L	1	0	0	0	0
12	M	1	0	0	0	0
All	All	29029	0	28496	1060	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 18.

The worst 5 of 1060 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:899:ILE:HD11	2:B:911:ILE:HA	1.23	1.10
6:H:81:PRO:HB2	6:H:82:PRO:HD3	1.29	1.07
1:A:567:LYS:HB2	1:A:568:PRO:CD	1.83	1.06
1:A:441:PRO:HD2	1:A:498:ARG:NH2	1.69	1.05
1:A:855:THR:HG21	1:A:857:ARG:HE	0.94	1.05

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	1398/1733 (81%)	1184 (85%)	154 (11%)	60 (4%)	3 35
2	B	1108/1224 (90%)	943 (85%)	105 (10%)	60 (5%)	2 30
3	C	264/318 (83%)	236 (89%)	20 (8%)	8 (3%)	5 46
4	E	212/215 (99%)	194 (92%)	18 (8%)	0	100 100
5	F	82/155 (53%)	74 (90%)	7 (8%)	1 (1%)	16 63
6	H	129/146 (88%)	98 (76%)	22 (17%)	9 (7%)	1 23
7	I	114/122 (93%)	97 (85%)	15 (13%)	2 (2%)	11 55
8	J	63/70 (90%)	56 (89%)	5 (8%)	2 (3%)	5 44
9	K	112/120 (93%)	106 (95%)	6 (5%)	0	100 100
10	L	44/70 (63%)	22 (50%)	14 (32%)	8 (18%)	0 3
11	M	183/345 (53%)	152 (83%)	19 (10%)	12 (7%)	1 25
All	All	3709/4518 (82%)	3162 (85%)	385 (10%)	162 (4%)	3 35

5 of 162 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	48	ALA
1	A	54	ASN
1	A	55	ASP
1	A	109	HIS
1	A	250	ILE

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	1224/1520 (80%)	1158 (95%)	66 (5%)	27 68
2	B	953/1061 (90%)	901 (94%)	52 (6%)	27 68
3	C	234/274 (85%)	225 (96%)	9 (4%)	40 76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	E	196/197 (100%)	192 (98%)	4 (2%)	63 87
5	F	74/137 (54%)	70 (95%)	4 (5%)	27 68
6	H	117/128 (91%)	110 (94%)	7 (6%)	24 65
7	I	113/116 (97%)	106 (94%)	7 (6%)	23 64
8	J	60/65 (92%)	57 (95%)	3 (5%)	30 70
9	K	99/102 (97%)	91 (92%)	8 (8%)	15 54
10	L	40/57 (70%)	34 (85%)	6 (15%)	3 26
All	All	3110/3657 (85%)	2944 (95%)	166 (5%)	28 69

5 of 166 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	408	LEU
2	B	764	SER
9	K	12	LEU
2	B	425	THR
2	B	547	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 64 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1432	GLN
2	B	516	ASN
7	I	90	GLN
2	B	46	GLN
2	B	300	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 9 ligands modelled in this entry, 9 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	1408/1733 (81%)	0.47	70 (4%) 32 21	82, 172, 267, 498	0
2	B	1122/1224 (91%)	0.85	157 (13%) 4 3	68, 183, 297, 489	0
3	C	266/318 (83%)	0.36	1 (0%) 93 87	93, 166, 244, 317	0
4	E	214/215 (99%)	0.48	20 (9%) 11 7	109, 202, 304, 500	0
5	F	84/155 (54%)	0.25	0 100 100	82, 135, 188, 256	0
6	H	133/146 (91%)	0.37	7 (5%) 30 20	118, 220, 305, 358	0
7	I	118/122 (96%)	1.52	36 (30%) 1 1	132, 237, 353, 412	0
8	J	65/70 (92%)	0.60	6 (9%) 11 7	106, 169, 237, 266	0
9	K	114/120 (95%)	0.25	0 100 100	94, 148, 208, 270	0
10	L	46/70 (65%)	0.64	6 (13%) 5 4	137, 208, 293, 359	0
11	M	187/345 (54%)	0.56	25 (13%) 4 4	111, 242, 373, 500	0
All	All	3757/4518 (83%)	0.60	328 (8%) 13 8	68, 181, 296, 500	0

The worst 5 of 328 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
11	M	121	LYS	15.7
7	I	119	THR	10.1
7	I	120	GLN	10.1
11	M	83	SER	9.3
7	I	53	GLY	9.2

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
12	ZN	M	346	1/1	0.89	0.50	3.14	193,193,193,193	0
12	ZN	C	319	1/1	0.98	0.31	1.16	191,191,191,191	0
12	ZN	A	1734	1/1	0.98	0.24	0.12	205,205,205,205	0
12	ZN	I	204	1/1	0.94	0.18	0.05	311,311,311,311	0
12	ZN	L	105	1/1	0.98	0.18	0.02	215,215,215,215	0
12	ZN	J	101	1/1	1.00	0.22	-0.72	177,177,177,177	0
12	ZN	A	1735	1/1	0.99	0.20	-0.73	183,183,183,183	0
12	ZN	B	1307	1/1	0.88	0.27	-	195,195,195,195	0
12	ZN	I	203	1/1	0.98	0.25	-	218,218,218,218	0

6.5 Other polymers (i)

There are no such residues in this entry.